时频系统故障预测*

王玲玲[1,2] 苟伟[1]
(1.中国科学院上海天文台，上海，200030;
2.上海市空间导航与定位技术重点实验室，上海，200030)

摘 要： VLBI 观测站的氢原子钟和时间比对系统（简称时频系统）为测站提供稳定可靠的时间和频率基准，系统信号的质量将直接影响观测数据质量。其中，氢原子钟是时频系统的关键设备，一旦出现故障，其对数据的破坏是致命的。因此，VLBI 站时频系统在设计建设之初就对氢原子钟及其周边各参数状态进行了实时监视并进行记录保存，一旦发现异常，技术人员需进行快速处理。中国 VLBI 网（CVN）建成至今，目前共五个测站，每个测站的时频系统都配有两台或以上的氢原子钟，系统长期运行以来积累了大量数据，通过对这些数据的分析，我们可以研究氢原子钟状态变化、性能变化、及环境变化之间关系，然后建立故障预测模型从而对设备故障进行预测。本文描述创建时频系统数据仓库，利用数据挖掘技术预测系统故障的测试过程，从测试结果可知在本项目中可以用聚类模型进行很好的预测。

关键词： 时频系统；氢原子钟；故障预测；数据驱动

1. 引言

时频系统为 VLBI 测站提供稳定可靠的时间标准，其输出信号的质量将直接影响观测质量。系统包括了氢原子钟和其他时间比对设备，氢原子钟作为其中最关键的设备，是一种高精度的频率标准，但由于其微波谐振腔输出频率易受外界环境因素干扰，所以其性能也很容易受外界环境影响，因此针对整个时频系统状态和性能变化的研究目前主要集中在环境变化对氢原子钟性能的影响上面，对此，国内外的学者都做了若干工作[1-4]。研究结果表明在环境控制较差的情况下，氢原子钟的频率稳定度会有显著的下降。因此，对于环境状态的有效控制是确保氢原子钟乃至整个时频系统可靠运行的前提。

另外一方面，作为用户，针对氢原子钟的运行，除了提高设备本身的可靠性之外，对发生故障的提前预测也是减少故障发生及尽早处理故障的有效方法。因为，氢原子钟的运行是一个动态的过程，其中伴随着复杂的物理和化学的状态变化，其中一些变化最终导致了氢原子钟性能指标的变化甚至导致了故障的发生，观察和分析这种变化之间的关系有助于实现故障预测。这首先需要对设备状态进行监视，然后通过将相关数据进行分析处理，得出预测模型，从而达到精准预测设备状态的目的。

2. 数据收集

故障预测技术是一门多学科综合信息处理技术，是故障诊断技术之上的一种拓展和延伸[5]，是在设备正常工作的前提下，利用历史数据和相关技术，结合当前状态对可能将要发生的故障趋势进行预测的一个过程[6]。预测方法的分类大致可以分为三类：基于模型、基于数据驱动、基于统计可靠性[5]。中国 VLBI 网（China VLBI Network，简称 CVN）建立已有十多年的历史，运行至今积累了大量的数据，具有进行相关研究的数据基础。本文选用基于数据驱动的故障预测技术来进行研究。该技术近年来广受重视并取得快速发展，并成为重要研究热点[7]。

---

* 受上海市空间导航与定位技术重点实验室基金资助
作者简介：王玲玲，女，高级工程师，研究方向：天文技术与方法。Email：llwang@shao.ac.cn

VLBI 测站时频系统的主要设备包括氢原子钟、分频钟/台站钟、时间间隔计数器、切换器、倍

频器、频率/脉冲分配器、GPS/BD 时间服务器等，另外还有提供周边服务的设备，包括 UPS、恒温恒湿空调、室内数字温度计、室外气象站等用以提供不间断电源及环境监测设备运行的外环境状态。根据《实用型氢原子钟技术说明书》，VLBI 测站使用的主动型氢原子钟提供类似小型钟的状态监控系统[8]，用户可以通过其串口采集数据，利用数据分析钟状态，其他设备数据的采集也同样可以通过设备自带的串口或网络通讯协议接口并配合对应的采集软件来完成，也可以根据需要通过专用串口/网口转换设备将接口进行转换后采集，数据采集频率根据需求设定。

时频系统的监测数据包括了氢原子钟内部各个参数的状态、钟差、钟房环境温湿度、外部环境温湿度、气压、风向风速等，不同的数据来自不同的设备，比如氢钟状态来自氢钟，钟房温湿度来自钟房温湿度计等。这些数据为预测模型的研究提供了第一手的资料，使得研究具有可行性。只是，经过多年建设及持续维护和改造，各时频系统中运行过或正在运行的设备众多，其中光氢原子钟的型号就包括了 MHM2010、VCH-1003M、SHOM-3、SHOM-4 及 SHOM-A 等多种，具体的设备编号也有所变动，数据存储的内容和格式有不同程度的差异，而且存储方式及数据格式并不完全相同，有的是以数据库的形式存储，有的是以文件存储，想要进行数据分析，第一步的工作是要对数据进行整理，建成时频系统数据仓库。

3. 数据整理

典型的基于数据驱动的故障预测方法有人工神经网络、模糊系统和其他的计算智能方法[5]。本文采用一种基于数据挖掘思想的预测算法展开研究。

时频系统的主要设备如前所述，经过业务分析，得出时频系统数据仓库的概念模型设计如图 1 所示。

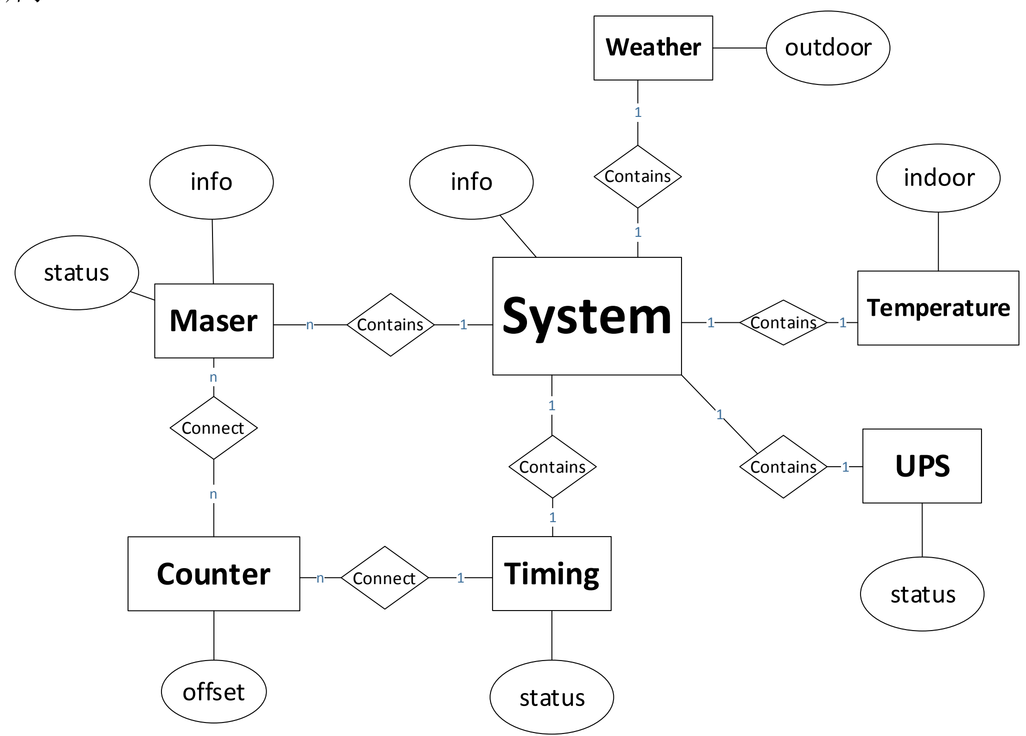


图 1 时频数据仓库的实体联系图

Figure 1 E-R diagram for timing system data warehouse

接下来，根据数据仓库建设的步骤，分别对其进行逻辑建模和物理建模，逻辑模型如图 2 所示，而部分数据表设计如表 1 ~ 表 6 所示。

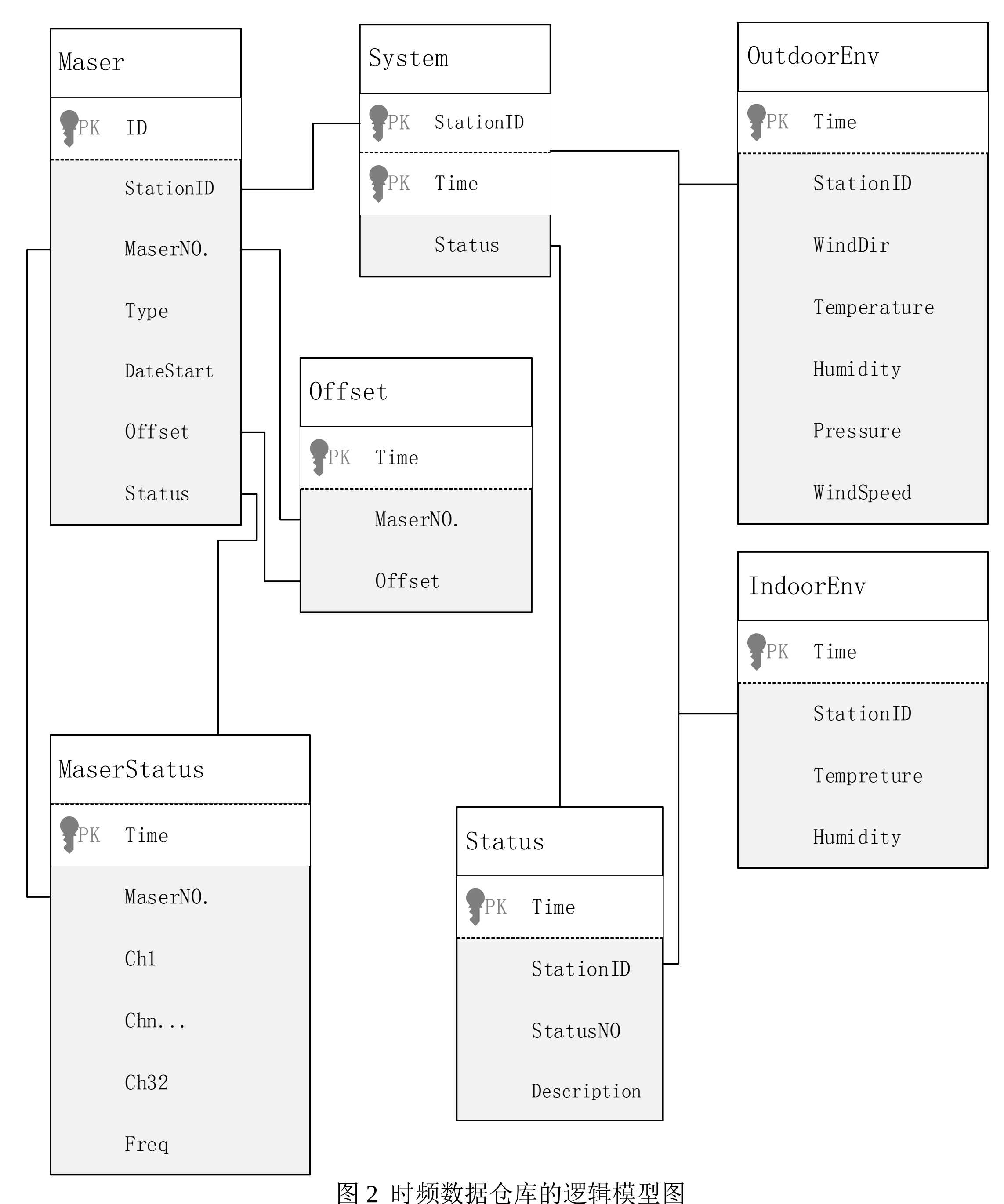


图 2 时频数据仓库的逻辑模型图
Figure 2 Logical model diagram for timing system data warehouse

表 1 氢钟信息
Table 1 Maser information

Column Name	Data Type	Allow Nulls
StationID	char(10)	N
MaserNO	nchar(7)	Y
Type	char(50)	Y
ID	char(20)	N pk
DateStart	datetime	Y

表 2 钟差数据
Table2 Clock offset

Column Name	Data Type	Allow Nulls
MaserNO	nchar(7)	N
Time	datetime	N pk
Offset	numeric(18,15)	Y

表 3 氢钟运行状态
Table 3 Maser status

Column Name	Data Type	Allow Nulls
MaserNO	nchar(7)	N
Time	datetime	N pk
Chan（1）	decimal(8,2)	N
Chan（…）	decimal(8,2)	N
Chan（32）	decimal(8,2)	N
Freq	nchar(20)	Y

表 4 室外环境
Table 4 Outdoor environment

Column Name	Data Type	Allow Nulls
StationID	char(10)	N
Time	datetime	N pk
Temperature	numeric(3,1)	N
Humidity	numeric(4,1)	N
Pressure	numeric(5,1)	N
WindDir	int	N
WindSpeed	numeric(3,1)	N

表 5 室内环境
Table 5 Indoor environment

Column Name	Data Type	Allow Nulls
StationID	char(10)	N
Time	datetime	N pk
Temperature	numeric(3,1)	N
Humidity	numeric(4,1)	N

表 6 系统运行状态
Table 6 System status

Column Name	Data Type	Allow Nulls
StationID	char(10)	N
Time	datetime	N pk
StatusNO	char(10)	Y
Descrip	char(100)	Y

由于数据的多样性和复杂性，首先需要对数据进行整理，整理的步骤包括数据清洗，及转换数据格式等。数据清洗即对原始数据中的错误数据以及无效数据进行剔除，再经格式转换后统一入库。

时频系统的数据记录有数据库、文本文件两种形式，清洗的过程中需要注意时间格式有北京时间和世界时的区别，为后续统一关系，这里全部换算成统一的世界时；台站的设备时有更换，有型号升级或同型号更换，这些信息有的完整，有的错误，有的缺失，缺失的补空，错误的需要修正；数据记录有混合，即同样的数据表格里，可能混合了不同时期不通型号或编号的设备数据，这种情况则需要根据当时的日志文件对数据进行拆分及重新补录；另外，由于源数据数据量较大，为方便测试，本文中对数据进行了均匀缩减抽样。

4. 数据分析

由于在所有数据中，能够最直观反映时频系统状态的是氢原子钟的状态数据，所以本文以 SHOM-4 型氢原子钟为例，重点分析有关氢原子钟的状态数据及当时的环境温度数据。根据该氢原子钟技术说明书的内容，钟的常规状态参数个数有 32 个，除去 1 个备用及 8 个隔离放大器的数据，其他均可以实时反映运行状态。图 3 是 2019 年编号为 88#的氢钟的状态及当时的环境

温度数据信息，其中纵坐标表示各参数的数值。

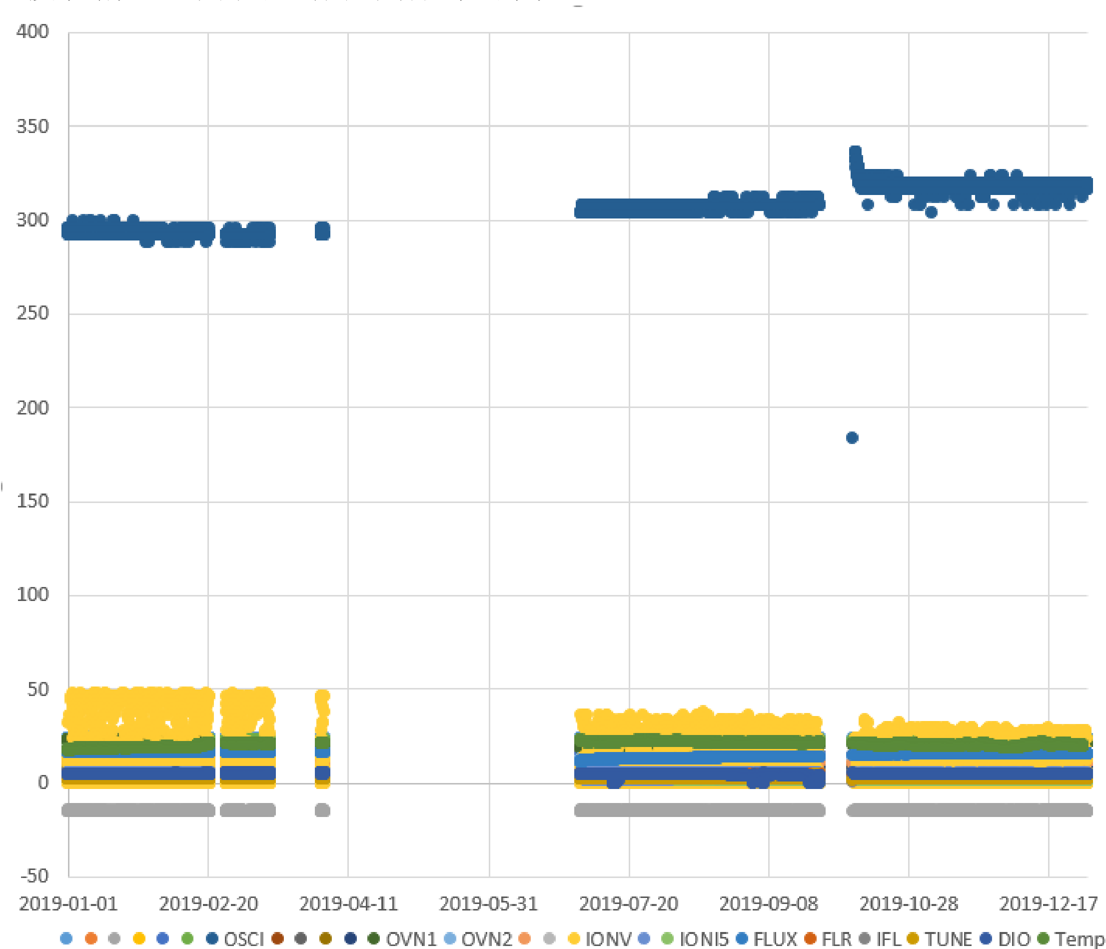


图 3 2019 年数据（部分数据有缺失）
Figure 3 Data in 2019 (some data are missing)

氢钟的这些状态参数可根据物理模块分为电源、恒温、离子泵、接收机、隔离放大器等几大部分，其中接收机中频信号 IFL 的数值代表钟信号，若值小于 1.5 则表示钟输出信号消失，属于严重故障，本文重点对该参数进行分析。在对数据进行细化之后，根据氢钟技术手册，再结合以往经验，可以得到共 11 个参数与 IFL 值密切相关，具体如图 4 所示，因为数据值相差比较大，为了让读者看得更清晰，所以把数据分成了两幅图显示，上半部分显示 OSCI 值，下半部分显示另外 10 个参数值。进一步细化，可以看到更多细节，比如从图 5 两个数据曲线形状来判断，认为两者有很强的相关性，可以说，曲线的弯曲方向大致是相反的，也就是温度升高时 OVN1 和 OVN2 值下降，温度下降时 OVN1 和 OVN2 值则升高，即 OVN1、OVN2 和温度数据相关度最高，这是因为 OVN1 和 OVN2 是钟恒温部分中外炉位置电压值，直接受环境温度变化影响。

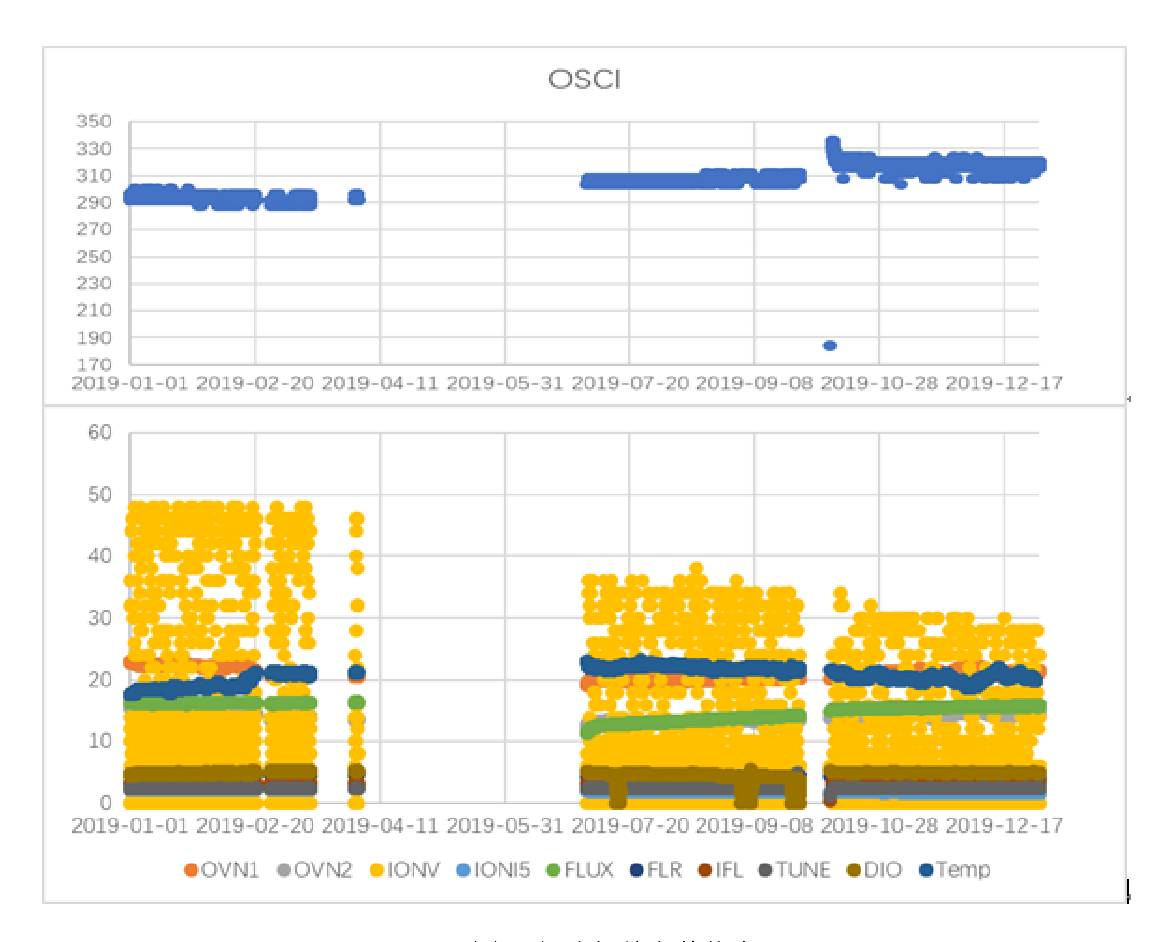


图 4 部分相关参数状态

Figure 4 Part of the relevant parameters status

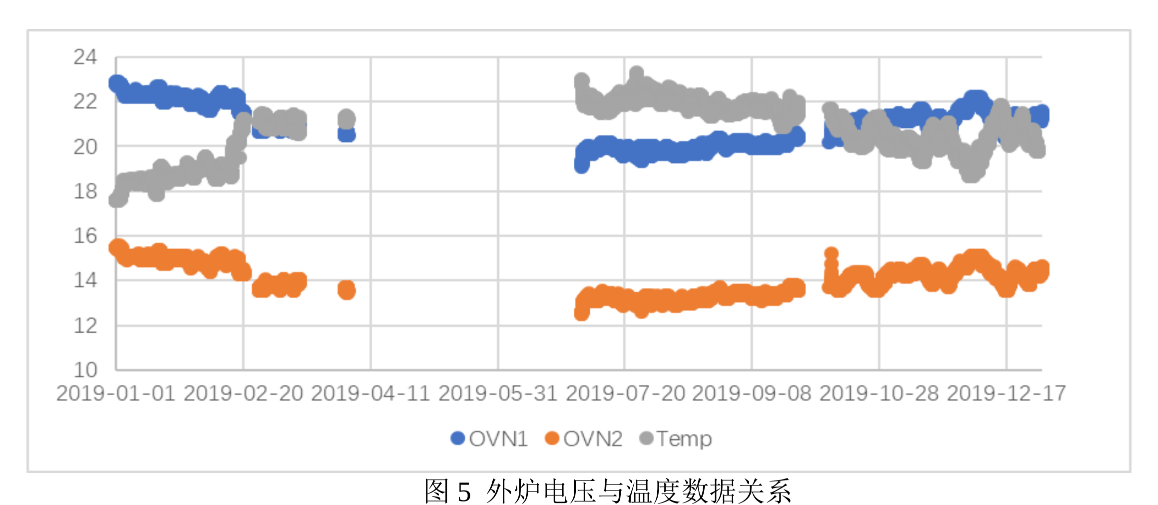


图 5 外炉电压与温度数据关系

Figure 5 the relationship between OVN1&2 and temperature

5. 故障预测模型

5.1. 几种挖掘算法的比较

信息化社会之下，日益增长的各类数据隐含着大量价值，数据挖掘技术是为了满足获取此价值的要求而来，是一种从大量随机数据中提取有用信息和知识的技术。该技术在发展过程中大量算法应运而生，但同时又各有各的局限，多种算法融合应该是解决具体事务的发展趋势[9]。本文以 2019 年数据为例，采用几种典型算法分别进行分析处理，并比较其性能。

根据每种算法在传统上的应用，以及结合本项目特定场景，选择了决策树、聚类、关联和时序等算法来进行测试研究。其中各算法的有关参数可以调整，比如，在决策树模型中，可以看到各参数的依赖关系网络，选择 IFL 节点，显示其所有链接，可以看到 IFL 与 OVN1、OVN2、FLUX、DIO 及温度数据有关，而调整到显示较强链接，可以看到与 DIO 的关系链断开了（图 6）；关联算法中多个参数由于是连续数据而被忽略；而时序算法需要连续的数据，不允许出现空白，因为测试中选定的数据中有缺失，所以需要指定用来填充空白的方法，可以使用前面的值、平均值或使用特定的常量。由各挖掘算法创建的挖掘模型如图 7 所示。

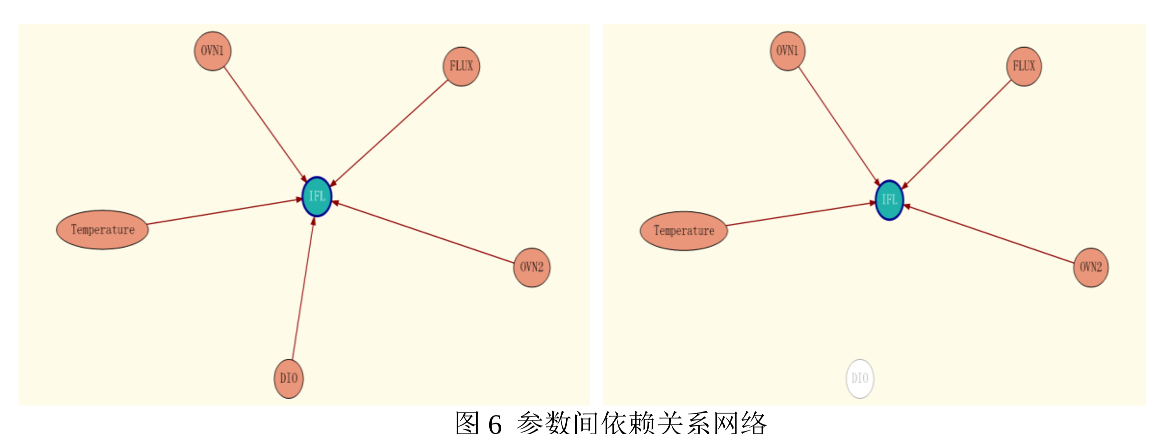


图 6 参数间依赖关系网络

Figure 6 Dependency network between parameters

结构	Tree	Cluster	Association
	Microsoft_Decision_Trees	Microsoft_Clustering	Microsoft_Association_Rules
DIO	Input	Input	忽略
FLR	Input	Input	Input
FLUX	Input	Input	忽略
IFL	PredictOnly	PredictOnly	PredictOnly
IONI	Input	Input	忽略
OSCI	Input	Input	忽略
OVN1	Input	Input	忽略
OVN2	Input	Input	忽略
Temperature	Input	Input	忽略
Time	Key	Key	Key

结构	TimeSeries
	Microsoft_Time_Series
DIO	Input
FLR	Input
FLUX	Input
IFL	Predict
IONI	Input
IONV	Input
OSCI	Input
OVN1	Input
OVN2	Input
Temperature	Input
Time	Key
TUNE	Input

图 7 各挖掘模型

Figure 7 Data mining models

可以使用提升图对各算法模型的准确性进行查看，如图 8 所示，蓝色线为理想模型，图中 X 轴表示用于比较预测的测试数据集的百分比，Y 轴表示准确预测的百分比。因此，理想线是一条对角线，用于显示在 50% 的数据中，模型准确预测 50% 的事例（即预计的最大值）。可见各模型都接近理想模型，其中决策树模型表现最为优秀。

然而，在仔细查看预测结果的时候，发现虽然绝大部分值都预测成功，但是数据中实际存

在的一个异常值并没有预测成功，即图 9 中鼠标选中的数值 0.38，因为小于 1.5，所以是实际的异常值，但是这个值在三个模型的预测中都没有被预测出来，原因可能是因为该异常值在 15130 个数据中只出现了一次，即测试数据不够典型，可想而知在实际预测中的预测效果并不好，时序算法中的测试存在同样的问题，所以接下来的测试对测试数据进行了重新选择。

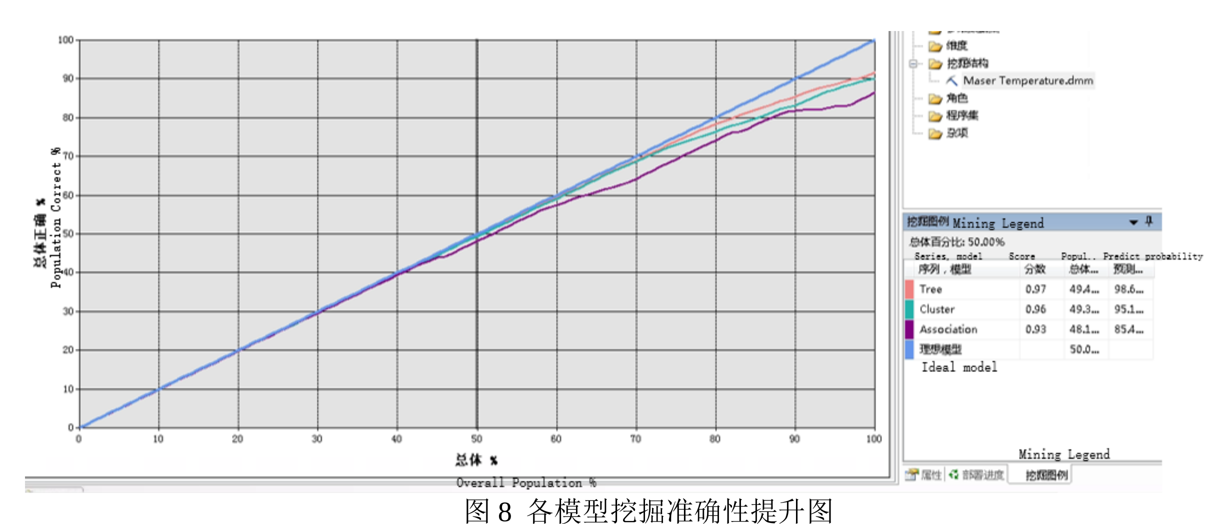


图 8 各模型挖掘准确性提升图

Figure 8 Data mining accuracy lift chart of the models

[IFL] 上 Tree 的计数:

预测 Predicted	3.12（实际） Actual	3.06（实际） Actual	3.08（实际） Actual	3.1（实际） Actual	0.38（实际） Actual
3.12	0	0	0	0	0
3.06	0	0	0	0	0
3.08	0	5	903	122	0
3.1	6	1	1142	12950	1
0.38	0	0	0	0	0

[IFL] 上 Cluster 的计数:

预测 Predicted	3.12（实际） Actual	3.06（实际） Actual	3.08（实际） Actual	3.1（实际） Actual	0.38（实际） Actual
3.12	0	0	0	0	0
3.06	0	0	0	0	0
3.08	0	5	1149	584	0
3.1	6	1	896	12488	1
0.38	0	0	0	0	0

[IFL] 上 Association 的计数:

预测 Predicted	3.12（实际） Actual	3.06（实际） Actual	3.08（实际） Actual	3.1（实际） Actual	0.38（实际） Actual
3.12	0	0	0	0	0
3.06	0	0	0	0	0
3.08	0	0	0	0	0
3.1	6	6	2045	13072	1
0.38	0	0	0	0	0

图 9 各模型的分类矩阵

Figure 9 Classification matrix of the models

5.2. 模型优化及预测

新的测试数据选自 2013 年的编号 43#氢钟，其中 IFL 的数据如图 10 所示，可以看出期间发生了多次信号消失现象，本文摘取这段数据进行模型建构展示。

这次模型建构测试中，聚类模型的分数高过决策树达到 0.88（图 11），所以选用该模型对测试数据的后半段进行预测，将预测数据中的预测值 IFL 置空（NULL）。从预测分类矩阵（图 12）中可以看出对于预测值 IFL，其各类数据的预测结果都比较均匀，与实际数据相符，预测结果中预测到值为 0.06（故障值）的数据 15 条，与实际数据的 14 条非常接近，预测概率从 0.51 到 0.857，准确率也很高。可知，用此模型进行故障预测成功。

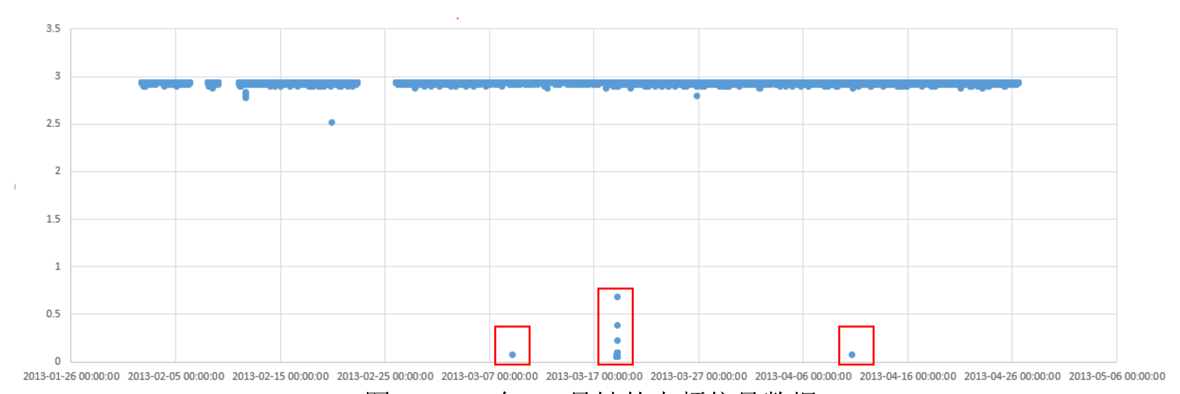


图 10 2013 年 43 号钟的中频信号数据

Figure 10 IFL data of 43# in 2013

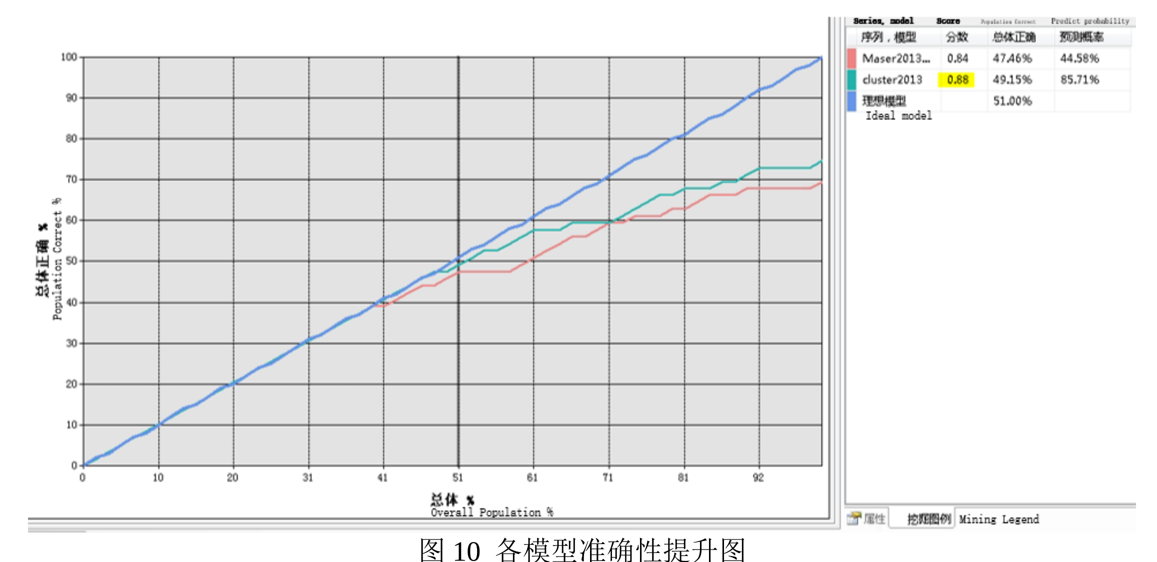


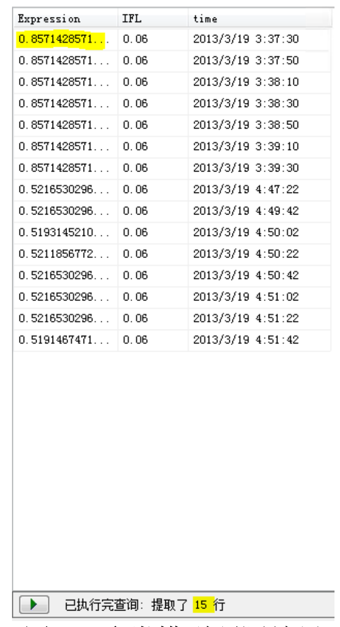
图 10 各模型准确性提升图

Figure 10 Data mining accuracy lift chart of the models

Predicted 预测	Actual 0.06 (实际)	Actual 0.08 (实际)	Actual 0.1 (实际)	Actual 0.38 (实际)	Actual 0.68 (实际)	Actual 2.92 (实际)	Actual 2.94 (实际)
0.06	10	4	0	1	0	0	0
0.08	4	5	1	0	1	0	0
0.1	0	0	0	0	0	0	0
0.38	0	0	0	0	0	0	0
0.68	0	0	0	0	0	0	0
2.92	0	0	0	0	0	4	3
2.94	0	0	0	0	0	1	25

图 11 聚类模型分类矩阵

Figure 11 Classification matrix of the clustering model



Expression	IFL	time
0.8571428571...	0.06	2013/3/19 3:37:30
0.8571428571...	0.06	2013/3/19 3:37:50
0.8571428571...	0.06	2013/3/19 3:38:10
0.8571428571...	0.06	2013/3/19 3:38:30
0.8571428571...	0.06	2013/3/19 3:38:50
0.8571428571...	0.06	2013/3/19 3:39:10
0.8571428571...	0.06	2013/3/19 3:39:30
0.5216530296...	0.06	2013/3/19 4:47:22
0.5216530296...	0.06	2013/3/19 4:49:42
0.5193145210...	0.06	2013/3/19 4:50:02
0.5211856772...	0.06	2013/3/19 4:50:22
0.5216530296...	0.06	2013/3/19 4:50:42
0.5216530296...	0.06	2013/3/19 4:51:02
0.5216530296...	0.06	2013/3/19 4:51:22
0.5191467471...	0.06	2013/3/19 4:51:42

已执行完查询：提取了 15 行

图 12 聚类模型预测结果

Figure 12 Prediction results of the clustering model

本项目同时对时序算法进行了测试。时序算法用于预测一系列连续数据点的未来值，要求必须包含时间键，而我们的数据正是以时间序列采集到的，符合要求。实际的预测结果如图 13 所示，可以看到左侧实线是实际值，右侧虚线是预测值，比对实际数据，即从图 10 中可以看到 3 月 19 日 5:17:10 之后附近的数据确实是在 3 左右，由此可知预测部分非常准确。然而，重新选择预测开始点的话，则预测的结果如图 14 所示，并不能令人满意，想来是因为数据的周期性不明显，因此使用的时序算法并不是最理想。这有待于后续继续分析研究。

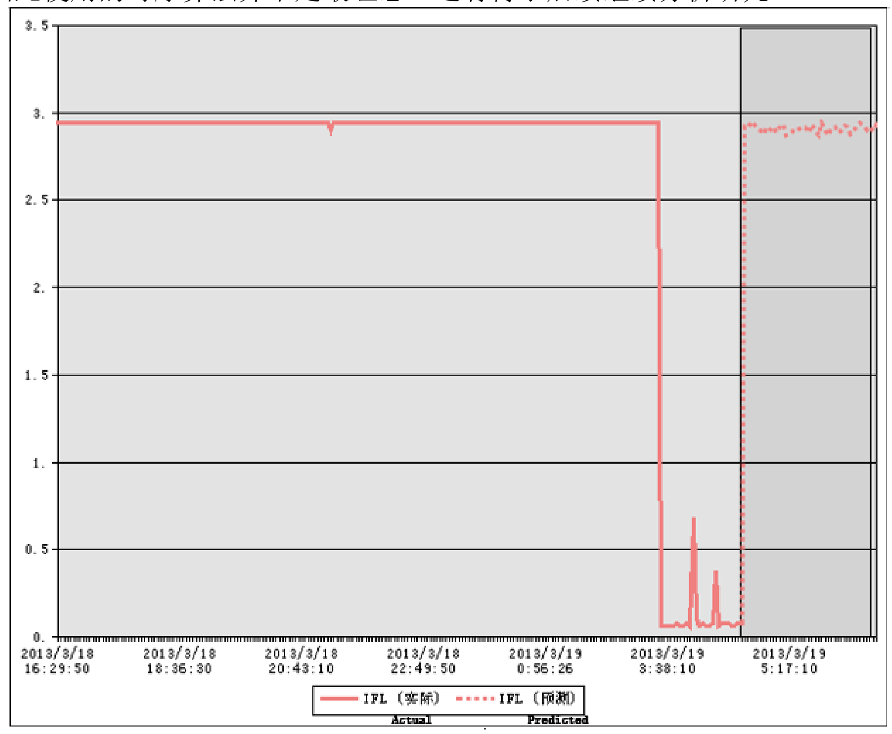


图 13 时序模型预测结果 1

Figure 13 Prediction results of the time series model

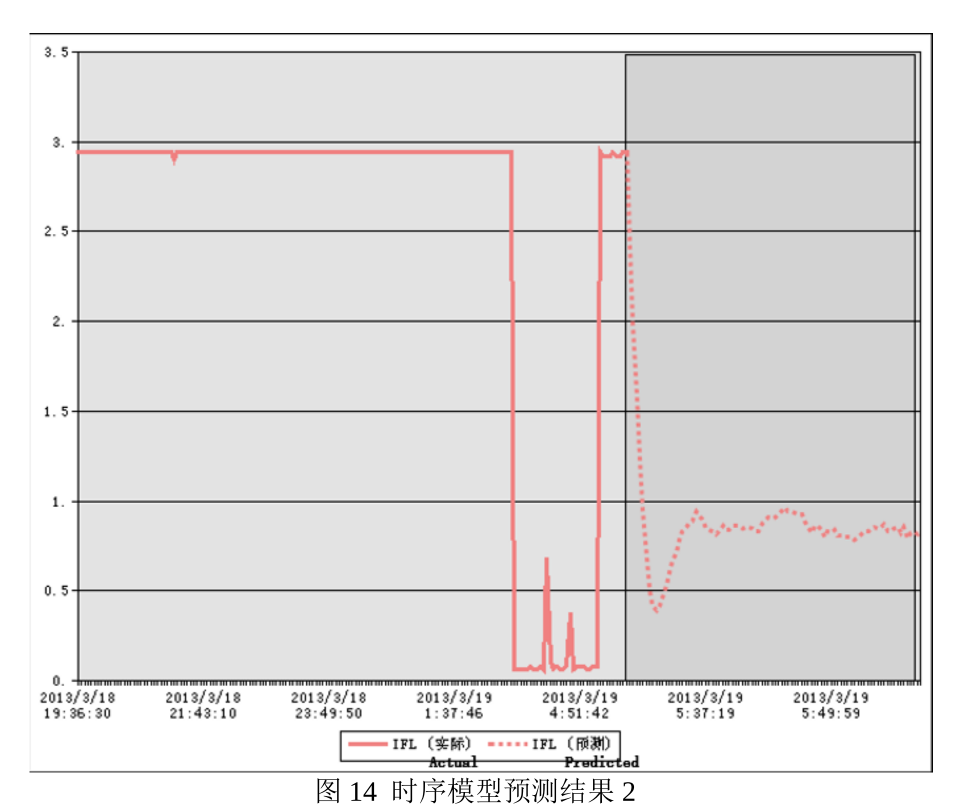


图 14 时序模型预测结果 2

Figure 14 Prediction results of the time series model

6. 总结

本文针对时频系统故障预测问题进行了一些方法探索，利用数据挖掘技术对历史数据进行了分析和测试。从测试过程可知，结合设备本身状态数据及运行环境数据可以分析预测设备状态趋势，从而对可能发生的故障进行预警。

分析的难点在于算法的选择和设计，模型建构的过程是算法不断优化的过程。本文测试得知聚类模型可很好地进行预测，后续可以通过更多的数据进行验证，同时下一步需将此预测模型集成到时频系统软件中从而实现系统实时故障预测。

另外，分析所用的预测模型需要有大量原始数据进行训练，用于训练的数据越多越全面越能覆盖所有可能则模型越准确越健壮，然而如何在有限的算力下高效处理这些数据成为难题，后续有必要进行专门研究。

Timing system failure prediction

WANG Lingling [1, 2] GOU Wei [1]
(1. Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai, 200030;
2. Shanghai Key Laboratory of Space Navigation and Positioning Technology, Shanghai, 200030)

Abstract: The hydrogen maser and time comparison system (timing system for short) of the VLBI observatory provides a stable and reliable time and frequency reference for the observing station. The quality of the system signal will directly affect the quality of the observation data. Among them, the hydrogen maser is the key equipment of the timing system. Once it fails, its damage to data is fatal. Therefore, at the beginning of the design and construction of the VLBI station timing system, the state of the hydrogen maser and its surrounding parameters were monitored and recorded in real time. Once an abnormality is found, the technicians need to deal with it quickly. China VLBI Network (CVN) has been established so far. There are currently five stations. The timing system of each station is equipped with two or more hydrogen masers. The system has accumulated a large amount of data since its long-term operation. Through the analysis of these data, We can study the relationship between state changes, performance changes, and environmental changes of the hydrogen maser, and then establish a failure prediction model to predict failures. This article describes the test process of creating a timing system data warehouse and using data mining technology to predict system failures, and we found the clustering model can be used to make good predictions in this project.

Key words: timing system; hydrogen maser; failure prediction; data-driven

参考文献:

[1] 杨军, 李世光, 徐月青, 等. 环境温度对氢原子钟性能的影响[J]. 测试技术学报, 2014, 28(2):93-97.

YANG Jun, LI Shiguang, XU Yueqing, etc. Effect of Temperature on Hydrogen Maser Performance[J]. Journal of Test and Measurement Technology, 2014, 28(2):93-97.

[2] 柯熙政, 刘宏良. 原子钟相位-时间起伏周期的小波分析[J]. 电子科技, 1997, 39(1):13-18.

Ke Xizheng, Liu Hongliang. Wavelet Analysis of Atomic Clock Phase-time Periodic Variations[J]. Electronic Science and Technology, 1997, 39(1):13-18.

[3] Richard A. Dragonette, Joseph J. Suter. Barometric pressure-induced frequency offs2ets in hydrogen masers[J]. Symposium on Frequency Control, 1991:586-590.

[4] 李树洲, 王茂磊, 肖胜红. 环境温度对氢原子钟稳定度的影响[J]. 现代导航, 2017, 8(2):118-121.

LI Shuzhou, WANG Maolei, XIAO Shenghong. Frequency Stability of Hydrogen Maser Affected by Environment Temperature[J]. Modern Navigation, 2017, 8(2):118-121.

[5] 彭宇, 刘大同, 彭喜元. 故障预测与健康管理技术综述[J]. 电子测量与仪器学报, 2010, 24(1):1-9.

Peng Yu, Liu Datong, Peng Xiyuan. A review: Prognostics and health management[J]. Journal of Electronic Measurement and Instrument, 2010, 24(1):1-9.

[6] 左宪章, 康健, 李浩, 等. 故障预测技术综述[J]. 火力与指挥控制, 2010, 1(35):1-5.

ZUO Xian-zhang, KANG Jian, LI Hao, etc. Overview of Fault Prediction Technology[J]. Fire Control & Command Control, 2010, 1(35):1-5.

[7] 彭宇,刘大同. 数据驱动故障预测和健康管理综述[J]. 仪器仪表学报, 2014, 3:481-495.

Peng Yu, Liu Datong. Data-driven prognostics and health management: A review of recent advances[J]. Chinese Journal of Scientific Instrument, 2014, 3:481-495.

[8] 陈丛军, 林传富. 小型氢原子钟监控系统的研制[J]. 中国科学院上海天文台年刊, 2007, 1:151-159.

CHEN Cong-jun, LIN Chuan-fu. The Development of The Monitor System for The Miniature Hydrogen Atomic Clock[J]. Annals of Shanghai Astronomical Observatory, 2007, 1:151-159.

[9] 邹志文, 朱金伟. 数据挖掘算法研究与综述[J]. 计算机工程与设计, 2005, 26(9):2304-2307.

ZOU Zhi-wen, ZHU Jin-wei. Research and summary of data mining algorithms[J]. Computer Engineering and Design, 2005, 26(9):2304-2307.